DOCUMENT RESUME

ED 343 954 TM 018 159

AUTHOR Woodruff, David; Houston, Mark
TITLE Reliability of Growth Rate in the Longitudinal Measurement of Gain.
PUB DATE 1 Apr 92
NOTE 6p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Gains; *Change; Equations (Mathematics); *Longitudinal Studies; Mathematical Models; *Measurement Techniques; Regression (Statistics); *Reliability
IDENTIFIERS Change Analysis; *Growth Curve Analysis; *Linear Models

ABSTRACT

The growth curve approach to the measurement of gain of D. Rogosa, D. Brandt, and M. Zimowski (1982) was more fully developed by J. B. Willet and D. Rogosa (1985). Willet later (1989) discussed some practical aspects of applying a linear growth model and reached the conclusion that reliability may not always be a useful index for the measurement of gain. This paper presents a detailed derivation of Willet's reliability formula and shows that his conclusion concerning the effect of the time intervals on the reliability of the growth rate is, somewhat paradoxically, not completely correct. It is concluded that the reliability of the estimated growth rate parameter should not be accorded much attention; rather, attention should be focused on the standard deviations of the estimated growth model parameters and Willet's (1989) advice should be followed so as to decrease these standard deviations. Two examples are given. A 6-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED343954

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

DAVID WOODRUFF

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

RELIABILITY OF GROWTH RATE IN THE LONGITUDINAL MEASUREMENT OF GAIN

David Woodruff & Mark Houston
American College Testing

TM018159

BEST COPY AVAILABLE

Reliability of Growth Rate in the Longitudinal Measurement of Gain

David Woodruff & Mark Houston—American College Testing—1 Apr 1992

Introduction

Gain scores and especially the reliability of gain scores has been a controversial topic in the measurement literature for over thirty years. Rogosa, Brandt, and Zimowski (1982) give a comprehensive and critical summary of the gain literature, and they argue persuasively that measuring gain is important and that gain can be well measured. They advocate a growth curve approach to the measurement of gain. The growth curve approach is more fully developed by Willet and Rogosa (1985) where it is discussed in the context of understanding correlates of change. Willet (1989) discusses some practical aspects of applying a linear growth model, and by critically examining his results this paper reaches the perhaps controversial conclusion that reliability may not always be a useful index for the measurement of gain.

Willet (1989) considers the longitudinal growth study situation where examinees are tested on two or more occasions at fixed time intervals with the goal of measuring change or rate of change over time. He assumes a linear model for individual growth and presents a formula for the reliability of the growth rate. He concludes from his formula that the spacing of the time intervals determine in part the reliability of the growth rate. This paper presents a detailed derivation of Willet's reliability formula, and it shows that his conclusion concerning the effect of the time intervals on the reliability of the growth rate is, somewhat paradoxically, not completely correct.

For a given individual, p, Willet (1989) assumes the following linear regression model for the observed measurement at time t_i:

$$X_{ip} = \xi_p(\bar{t}) + \theta_p(t_i - \bar{t}) + e_{ip} \quad i = 1, \cdots, T. \tag{1}$$

For each person, the conditional model may be viewed as a linear regression model of X_{ip} onto $(t_i - \bar{t})$ with T observations. Note that $\bar{t}$ is the mean of the $\{t_i\}_{i=1}^{T}$ times, and these time are the same for every individual. It is assumed that within each individual the measurement errors $\{e_i\}_{i=1}^{T}$ are uncorrelated and that the within person variance of these errors, σ_e^2, is the same for every individual.

The parameter $\xi_p(\bar{t})$ is the intercept in the regression of X_{ip} onto $(t_i - \bar{t})$, and it represents examinee p's predicted score at time $\bar{t}$. Because the within individual model is a linear regression model, regression theory yields $\overline{X}_{\cdot p}$, the mean of examinee p's T observed scores, as the unbiased minimum variance estimator for each individual's intercept parameter. The parameter θ_p is the regression slope of X_{ip} onto $(t_i - \bar{t})$; it is the most meaningful parameter for characterizing individual growth, and it is called the growth rate parameter (Rogosa, Brandt, and Zimowski, 1982). Regression theory yields the following unbiased minimum variance estimator of each individual's growth rate parameter:

$$\hat{\theta}_p = \frac{s(X_{ip},(t_i-\bar{t})|p)}{s^2((t_i-\bar{t})|p)} = \sum_{i=1}^{T} \frac{(t_i-\bar{t})X_{ip}}{\mathrm{SST}}, \tag{2}$$

where $\mathrm{SST} = \sum_{i=1}^{T} (t_i - \bar{t})^2$. Note that the covariance and variance in 2 are conditional for a fixed examinee, p, and the notation reflects this. Also note that $\hat{\theta}_p$ is a linear function of examinee p's observed scores, $\{X_{ip}\}_{i=1}^{T}$, because the $\{t_i\}_{i=1}^{T}$ are fixed constants. If $T = 2$, $t_1 = 0$, and $t_2 = 1$, then $\hat{\theta}_p = X_{2p} - X_{1p}$. Of course $\overline{X}_{\cdot p}$ is also a linear function of examinee p's T observed scores, and regression theory dictates that the estimated slope is uncorrelated with the estimated intercept. Henceforth, only the more informative slope parameter and its estimator will be considered in detail here.

TM018159

Letting $\sigma^2(\theta)$ denote the variance of θ_p over all examinees, Willet (1989) presents the following formula for the reliability of $\hat{\theta}$:

$$\rho(\hat{\theta}) = \frac{\sigma^2(\theta)}{(\sigma^2(\theta) + \sigma_e^2/\mathrm{SST})}, \tag{3}$$

and he notes that the reliability may be increased by increasing SST which is solely a function of the measurement times, and these are often under the control of the researcher. The following detailed derivation of the above reliability formula indicates that SST does not necessarily affect reliability estimation, despite its pivotal role in the estimation of each individual's growth rate parameter.

Derivation

Let the T by 1 column vector $\mathbf{a}$ have elements $\{a_i = t_i - \bar{t}\}_{i=1}^{T}$, and let $\mathbf{b} = \mathbf{a}/\mathrm{SST}$. Note for future reference that $\mathbf{b}'\mathbf{b} = 1/\mathrm{SST}$. Let the column vector of the T observed scores for examinee p be denoted by $\mathbf{x}_p$ and the vector of T true scores by τ_p where $\tau_{ip} = \xi_p(\bar{t}) + \theta_p(t_i - \bar{t})$. With this new notation

$$\hat{\theta}_p = \mathbf{b}'\mathbf{x}_p \quad \text{and} \quad \theta_p = \mathbf{b}'\tau_p . \tag{4}$$

By assumption the T by T conditional covariance matrix for $\mathbf{x}$ given p is $\Sigma(\mathbf{x}|p) = \sigma_e^2\mathbf{I}$. Necessary for the derivation of the reliability formula is the T by T unconditional covariance matrix, $\Sigma(\tau)$, which is the covariance matrix of true scores on intra-individual repeated measurements over the population of all examinees. It is very likely, if not essential, that examinee true scores obtained from multiple measurements over time be correlated and hence that $\Sigma(\tau)$ not be diagonal. As will be seen later, however, the form of $\Sigma(\tau)$ is not critical to the main result so that no assumptions regarding its form need be made.

Lord and Novick (1968, p 61) define reliability as the squared correlation between observed score and true score. In the present application this leads to

$$\rho(\hat{\theta}) = \rho^2(\theta,\hat{\theta}) = \frac{\left[\sigma(\theta,\hat{\theta})\right]^2}{\sigma^2(\theta)\sigma^2(\hat{\theta})} . \tag{5}$$

In what follows it should be noted that though θ is "tied" to p, the values of p are unique, whereas two examinees may have the same value of θ. Letting μ denote expectation, using results from Rao (1973, p 97 & p 153), and substituting from 4 yields the following results:

$$\sigma(\theta,\hat{\theta}) = \sigma(\theta,\mu[\mu\{\hat{\theta}|p\}|\theta]) = \sigma^2(\theta) = \sigma^2(\mathbf{b}'\tau) = \mathbf{b}'\Sigma(\tau)\mathbf{b} , \tag{6}$$

$$\sigma^2(\hat{\theta}) = \sigma^2(\mu[\hat{\theta}|p]) + \mu(\sigma^2[\hat{\theta}|p]) = \mathbf{b}'\Sigma(\tau)\mathbf{b} + \mathbf{b}'\mathbf{I}\mathbf{b}\sigma_e^2 , \text{ and} \tag{7}$$

$$\sigma^2(\theta) = \sigma^2(\mu[\theta|p]) + \mu(\sigma^2[\theta|p]) = \mathbf{b}'\Sigma(\tau)\mathbf{b} , \tag{8}$$

the last result deriving from the fact that $\sigma^2(\theta|p) = 0$ for all p. Substituting these results into 5 and recalling that $\mathbf{b}'\mathbf{b} = 1/\mathrm{SST}$ yields:

$$\hat{\rho}(\theta) = \frac{\sigma^2(\theta)}{\sigma^2(\theta) + \left(\sigma_\varepsilon^2/\mathrm{SST}\right)} = \frac{\mathbf{b}'\Sigma(\tau)\mathbf{b}}{\mathbf{b}'\Sigma(\tau)\mathbf{b} + \mathbf{b}'\mathbf{I}\mathbf{b}\sigma_\varepsilon^2} . \tag{9}$$

The last expression for $\hat{\rho}(\theta)$ shows that all terms in both the numerator and the denominator are quadratic forms in **b**, and that changing **b** affects all terms in a similar manner. Changing **b** to increase SST will not necessarily increase the reliability of the growth rate.

Two Examples

Consider the example for which there are only two measurements for each examinee—one at time t_1 and the other at time t_2. Here $\bar{t} = (t_1 + t_2)/2$, $\mathrm{SST}\mathbf{b}' = [(t_1 - \bar{t}), (t_2 - \bar{t})]$, and $\Sigma(\tau)$ has diagonal elements σ_1^2 and σ_2^2 and off-diagonal element σ_{12}. Substituting the above into 9 and then simplifying yields:

$$\hat{\rho}(\theta) = \frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12} + 2\sigma_\varepsilon^2} , \tag{10}$$

which is totally independent of t_1 and t_2 and hence of SST. The covariance σ_{12} might decrease as the interval between the two measurement—and hence SST—increased, but this would not necessarily increase the reliability of the growth rate because both numerator and denominator would be affected.

For another example take the true score variance for each repeated measurement as having the same value, σ_τ^2, and the covariances among the T repeated measurements as homogeneous with value γ_τ. Then

$$\Sigma(\tau) = \sigma_\tau^2\mathbf{I} + (\mathbf{J} - \mathbf{I})\gamma_\tau , \tag{11}$$

where **J** is a T by T matrix with every element equal to unity, and

$$\hat{\rho}(\theta) = \frac{(\sigma_\tau^2 - \gamma_\tau)/\mathrm{SST} + \mathbf{b}'\mathbf{J}\mathbf{b}}{(\sigma_\tau^2 - \gamma_\tau + \sigma_\varepsilon^2)/\mathrm{SST} + \mathbf{b}'\mathbf{J}\mathbf{b}} = \frac{(\sigma_\tau^2 - \gamma_\tau)}{(\sigma_\tau^2 - \gamma_\tau + \sigma_\varepsilon^2)} \tag{12}$$

because $\mathbf{b}'\mathbf{J}\mathbf{b} = 0$. Here again, the reliability of the growth rate is unaffected by SST.

Discussion

At first sight, Willet's (1989) result appears puzzling because the time values are constants and constants should not affect the reliability of any parameter estimator. However, regression theory shows that the standard deviation of the estimator of an individual's growth rate parameter has SST in its denominator and that increasing SST either by increasing T (it can be shown that SST is a non-decreasing function of T) or just rearranging the measurement times with T fixed reduces this standard deviation. Furthermore, increasing T also decreases the standard deviation of the estimator of an individual's intercept. Hence, one can obtain more accurate estimates of the parameters in each individual's linear growth model by following the suggestions given by Willet (1989), but, somewhat paradoxically, his suggestions do not necessarily increase the reliability of the estimated growth rate because reliability is the ratio of true score variance over observed score variance, and, as shown above, SST cancels itself out because it is in the denominator of both variances.

The above suggests that the reliability of the estimated growth rate parameter should not be accorded much attention, but rather attention should be focused on the standard deviations of the estimated growth model parameters, and Willet's (1989) advice followed so as to decrease these standard deviations. Willet (1989) notes that an additional advantage of increasing T to at least 3 is that the fit of the linear model can then be investigated. If there is doubt about the adequacy of a linear model for growth in a particular situation, then at least 3 measurements per examinee are essential. This paper assumes a linear growth model, and is mute on the appropriateness of such a model for a particular situation, except to suggest that if the number of measurements per examinee must be limited to two and there is some uncertainty about the adequacy of a linear growth model, then Draper and Smith (1981, p 25) should be consulted for advice on how best to choose the time points.

References

Draper, N. R. & Smith, H. (1981). *Applied Regression Analysis*. New York: Wiley.

Lord, F. M. & Novick, M. R. (1968) *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Rogosa, D., Brandt, D. & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92*, 726-748.

Rao, C. R. (1973). *Linear Statistical inference and its applications*, (2nd ed.). New York: Wiley.

Willet, J. B. & Rogosa, D. R. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*, 203-228.

Willet, J. B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement, 49*, 587-602.